Defect Initiation, Growth, and Failure –
A General Statistical Model and Data Analyses

Wayne B. Nelson
Consultant, Schenectady, NY, WNconsult@aol.com

Abstract

This paper describes a simple and versatile new model for defect initiation and growth leading to specimen failure. This model readily yields the distribution of time to failure (when a specified defect size is reached). The model can be readily fitted to growth and failure data, using existing software for fitting regression models to censored data. The model and fitting methods are illustrated with applications to dendrite growth on circuit boards and growth of cracks in catalytic converter containers.

1. Introduction

In some applications, a product is observed over time. Initially it is defect free. After a random time, a defect initiates and grows stochastically over time. In some applications, the product fails when the defect reaches a certain size. When there are multiple defects, the progress of the largest defect is followed. This article presents a new model and fitting methods for such defect initiation and growth.

Examples of applications are:

- Dendrites initiate and grow between two parallel conductors on a circuit board. When a dendrite grows across the gap, it causes a short and a circuit malfunction.
- Cracks initiate and grow in a noncritical component of a jet engine. When a crack reaches a specified size, the component has defined "failure" and is replaced.
- Cracks initiate and grow in the container wall for a catalytic converter on cars. The wall fails when a piece falls out and beads of catalyst escape.

In such applications, one needs a model to understand, predict, and manage such defect initiation, growth, and failure, say, by replacing a unit at a specified age or defect size. In particular, one seeks

- to estimate model parameters and functions of them,
- to estimate the cumulative distribution $G(t)$ of time to defect initiation, and
- to estimate the cumulative distribution $H(t)$ of time to reach a specified defect size (resulting in a defined or a catastrophic failure).

Section 2 describes a general model, which provides such information. This new degradation model does not appear in Meeker and Escobar (1998) or Nelson (1990), which together survey basic statistical models and theory for product degradation. Section 3 describes how to fit the model to such data using computer packages that handle censored data.

2. A Population Model for Defect Initiation and Growth

This section describes a population model for such data.

The positive size $Y_i(t)$ of the defect on population unit i as a function of time t is an unspecified stochastic process. In most applications, $Y_i(t) = 0$ up to the initiation time T_i and then increases; for example, cracks do not get shorter.

At any age t, there is a population cumulative distribution function $F(y;t)$ for defect size Y. While size Y is usually physically positive, $F(y;t)$ is defined for $-\infty < y < \infty$ for reasons explained below. Then $F(0;t)$ is interpreted as the fraction of the population that is has not initiated a defect by age t. Consequently, $G(t) = 1 - F(0;t)$ is the cdf of time to crack initiation. Similarly, if a unit "fails" when its defect size reaches Y_0, the population fraction with size below Y_0 is $F(Y_0;t)$, and $H(t) = 1 - F(Y_0;t)$ is the cdf of time t to failure.

For convenience, suppose that the distribution of size Y at age t has a location parameter $\lambda(t)$ and a scale parameter $\sigma(t)$, which depend on t. These regression functions might be based on physical theory for defect growth. For concreteness, suppose that

$$\lambda(t) = \beta_0 + \beta_1 t \quad \text{and} \quad \sigma(t) = \beta_3 ,$$

where the β_j coefficients are to be estimated from data. Also, for concreteness, imagine that $F(y;t)$ is a normal cdf and $\lambda(t)$ and $\sigma(t)$ are its median and standard deviation. That is, $F(y;t) = \Phi\{[y-\lambda(t)]/\sigma(t)\}$, where $\Phi\{\ \}$ is the standard normal cdf. Of course, other distributions and parameter functions can be used. In particular, $\sigma(t)$ may be a function of t and coefficients.

3. Model Fitting by Computer

Software for fitting regression models to censored data can easily be used to fit the preceding model to simple defect growth data where, for unit i, the size Y_i of the defect is observed only once at some age t_i. That is, the unit is not observed repeatedly over time. This was so for the applications above; indeed, inspection of the circuit destroys it. With one observation per unit, the autocorrelation structure of the $Y_i(t)$ process cannot and need not be modeled.

Such data consist of n units where some have a defect and others do not. Then data on those with a defect can be expressed as (Y_1,t_1), (Y_2,t_2), $\cdots$, (Y_r,t_r), where r is the random number of units with a defect. Data on those without a defect can be expressed as $(0,t_{r+1})$, $(0,t_{r+2})$, $\cdots$, $(0,t_n)$.

For purposes of fitting the model above to such data, each unit with a defect is treated as an observed value (Y_i,t_i). Each unit without a defect is treated as censored on the left at 0, that is, as the value $(0-,t_i)$. Such fitting is done with the method of maximum likelihood (ML) by various software packages. They provide estimates of model parameters and functions of them, such as $G(t)$ and $H(t)$, and corresponding approximate confidence limits.

The ML fitting method depends on the usual assumptions, such as:

1) The n observed units are statistically independent, that is, are a simple random sample.

2) The observed Y_i values are statistically independent of their observation times t_i.

One could directly estimate the distribution of time to defect initiation as follows. The data above provides quantal-response data on initiation times. There t_1, t_2, $\cdots$, t_r are left censored initiation times, and t_{r+1}, t_{r+2}, ···, t_n are right censored initiation times. Software will fit a distribution to such data. Similarly, one could directly estimate the distribution of time to failure. Then there are two situations:

1) The q failures are catastrophic when $Y(t)$ reaches Y_0, and their exactly observed times are t_1, t_2, ···, t_q. The remaining times t_{q+1}, t_{q+2}, ···, t_n are right censored times. The dendrite failures are catastrophic. Software will fit a distribution to such data.

2) The failures are not catastrophic and $Y(t)$ values may exceed Y_0. The jet engine component cracks are of this type. Then t_i values corresponding to Y_i values greater than Y_0 are left censored times from the failure time distribution. Also, the remaining t_i values are right censored. Software will fit a distribution to such quantal-response data.

Of course, these three estimates ignore the size values Y_1, Y_2, ···, Y_r, and each is less accurate than the fitted model, provided it is correct.

Such a fitted regression model yields censored residuals, which can be analyzed as usual to assess the well the model fits the data and to check for peculiar data, as described by Nelson (1990). Also, likelihood ratio tests can be used to compare the fit of various models.

In conclusion, this simple model and the fitting method are useful and easy in practice. Suitable physical models need to be developed for applications. Also, optimum or good test plans (choice of inspection times) need to be developed.

References

Meeker, W.Q. and Escobar, L.A. (1998), *STATISTICAL METHODS FOR RELIABILITY DATA*, Wiley, New York.

Nelson, Wayne (1990), *ACCELERATED TESTING: STATISTICAL MODELS, TEST PLANS, AND DATA ANALYSES*, Wiley, New York. Paperback edition (2004).